Universite Jean Monnet

Data Mining Project with R

Master 1 MLDM

Applying Data Mining Techniques on Global Terrorism Dataset

Author:
Gulsah Samut

Supervisor:
Fabrice Muhlenbach

March 12, 2018

Introduction

According to a survey, about 218 million people are affected by calamities, natural and man-made, per annum and about 68000 people loose their lives every year. The frequency of natural disasters like earthquakes, volcanoes, etc have remained broadly constant, but the number of terrorist activities have grown over the period. The aim of this project is to explore the terrorist events around the world. Interactive Plots and Animations are used in this project, for making the exploration easy and more informative.

1. Problem Understanding

This dataset contains very much information about the terrorism, from 1970 to 2016. There are 170350 records and 135 columns including date, time, location, number of hostages, killed, wounded, if there was a ransom, the outcome, if there was a suicide attack, claims, weapons used.

I have analyzed as well the perpetratos group and the perpetrators nationality, the geographical region where the attacks occured, the location of escape of the perpetrators (the country or city they asked to go after the attack), the weapons used in the attack, the claim mode (channel used to transmit their claim), the property damages. Looking to the specific terrorist groups activity, I also correlate those groups activity with certain political measures in the countries where the groups are acting.

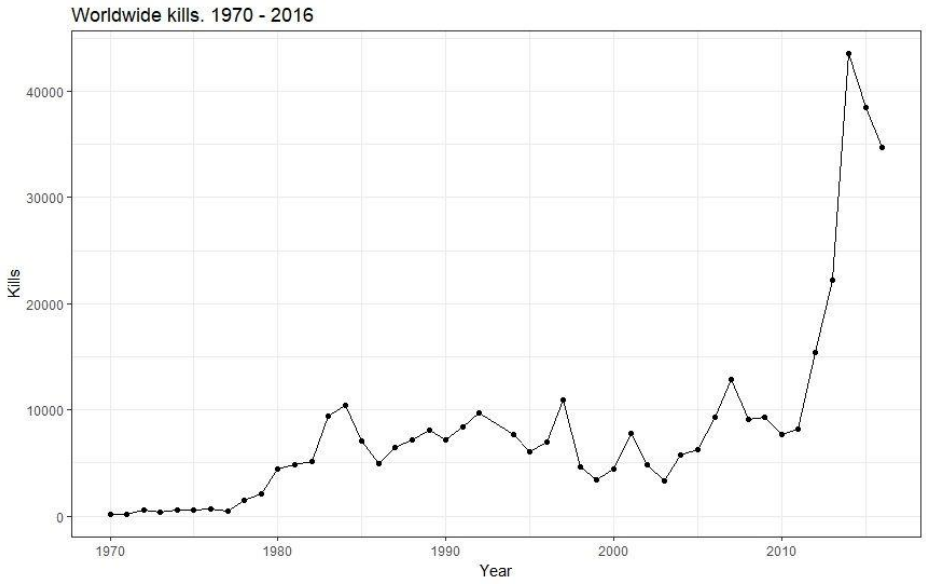
2. Data Understanding

Characteristics of the GTD

- Contains information on over 170,350 terrorist attacks
- Currently the most comprehensive unclassified database on terrorist events in the world
- Includes information on more than 83,000 bombings, 18,000 assassinations, and 11,000 kidnappings since 1970
- Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables
- Over 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2016 alone

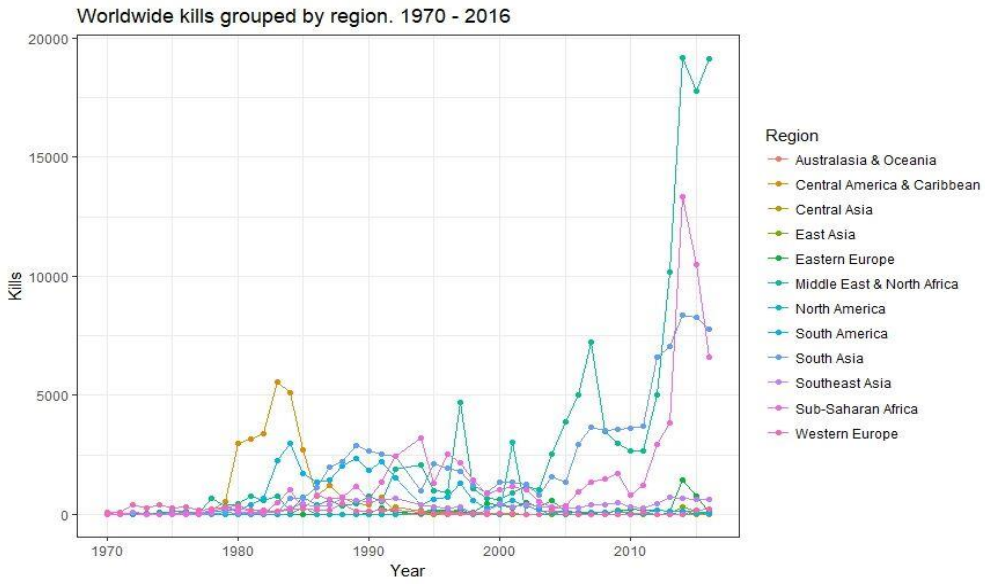
This dataset has been analyzed according to Worldwide Kills between 1970-2016, Worldwide Kills Group by Region between 1970-2016, Number of Attacks by Region between 1970-2016, Top 15 Countries with Highest Number of Attacks between 1970-2016, Top 10 Countries in Europe with Highest Number of Attacks between 1970-2016, Worldwide Attacks with Kills between 2000-2016, Number of Attacks by Type of Weapon, Number of Attacks by Target and Number of Attacks by Attack Type.

Worldwide Kills between 1970-2016



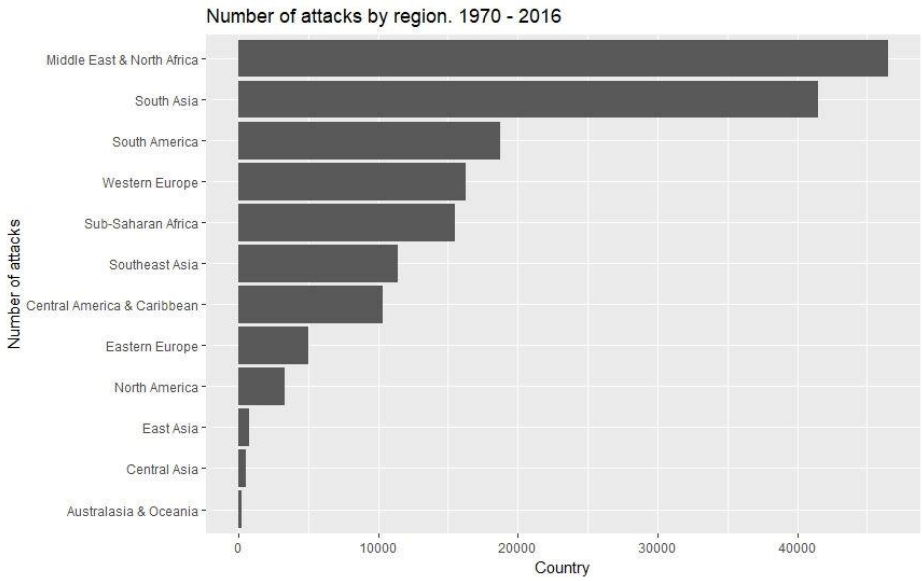
This graph shows that number of worldwide kills was stable between the years 1970 and 1980. After 1980 numbers started to increase until 1985 and it reached 10000. Worldwide kills fluctuated between increaing and decreasing until the year of 2000.We can easily see from this graph that the number of the worldwide kills have dramatically increased after 2000.

Worldwide Kills Group by Region between 1970-2016



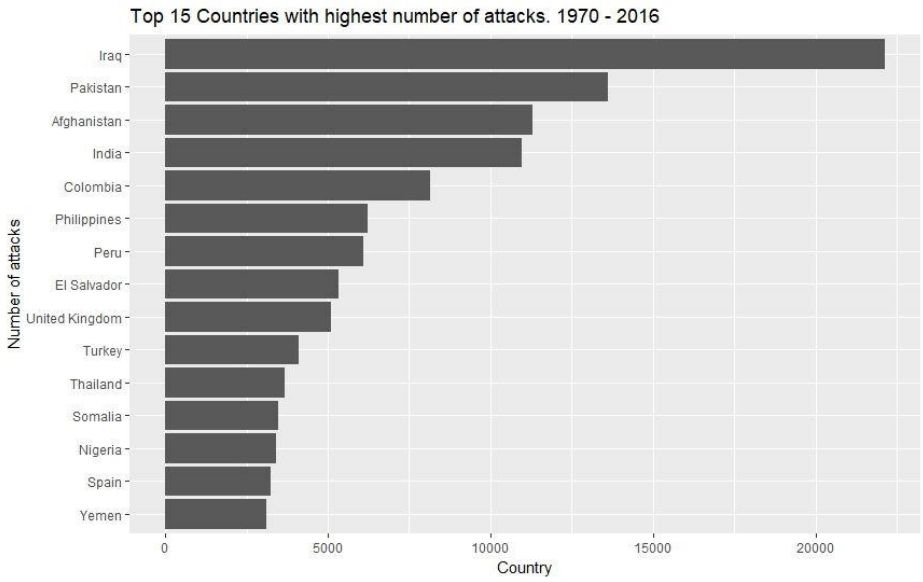
This graphs shows the worldwide kills grouped by region between 1970-2016. As it can be seen regions like Central Asia, Australiasia & Oceania, East Asia did not affected by terrorism between 1970-2016 as much as the other regions did. From 1978 to 1982, the number of kills had increased in the region of Central America & Caribbeans but after that time forth, this regions have not affected by terrorism much. This graphs shows that Middle East & North Africa has been highly affected by terrorism.Number of kills reached 20000 after the year of 2010 in this region.

Number of Attacks by Region between 1970-2016



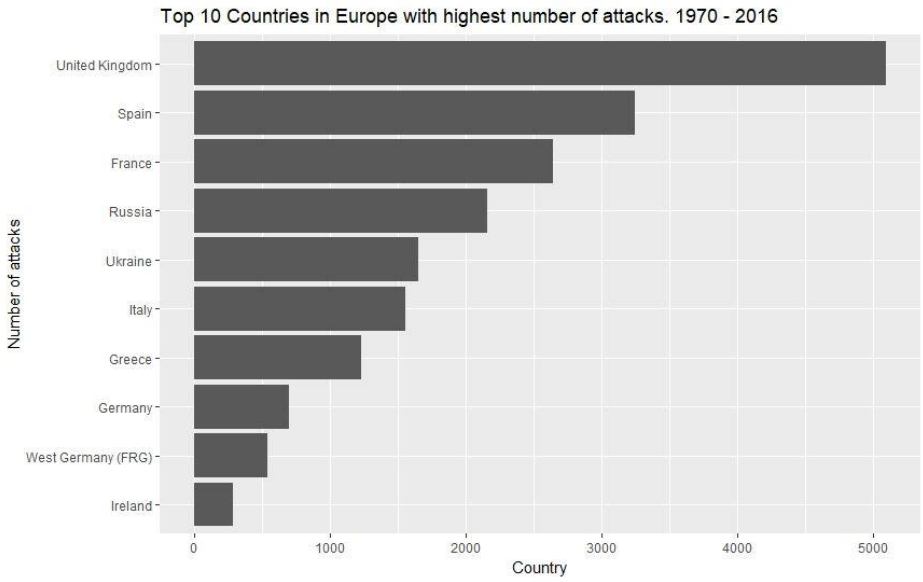
This graph shows number of attacks by region between 1970- 2016. It can easily seen Middle East & North Africa is coming first by the number of attacks. This number reached more than 40000 by the year of 2010. South Asia is in the second place fort he number of attacks. Sout America, Western Europe, Sub-Saharan Africa also have high number of attacks. Australasia & Oceania is in the last place fort he number of attacks. The number of attacks are almost negligible.

Top 15 Countries with Highest Number of Attacks between 1970-2016



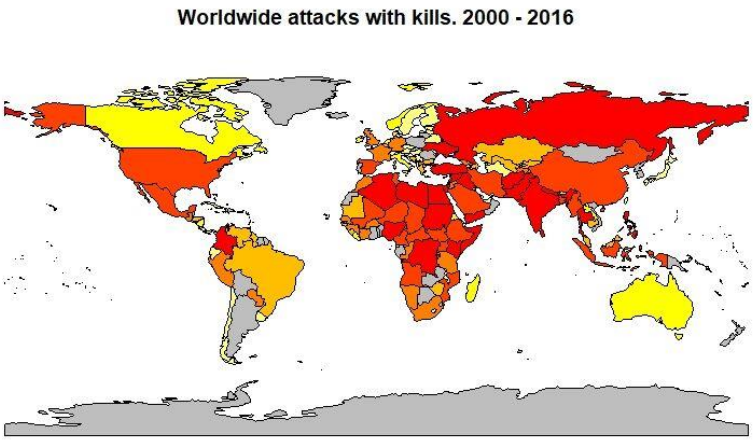
This graph shows that Iraq is coming first fort he number of attacks that happened between 1970-2016. Between these years, number of attacks reach more than 20000 in Iraq which is located in Middle East region. Pakistan is following Iraq and it proves that South Asia comes in second.Number of attacks in Pakistan reached more than 12000.

Top 10 Countries in Europe with Highest Number of Attacks between 1970-2016



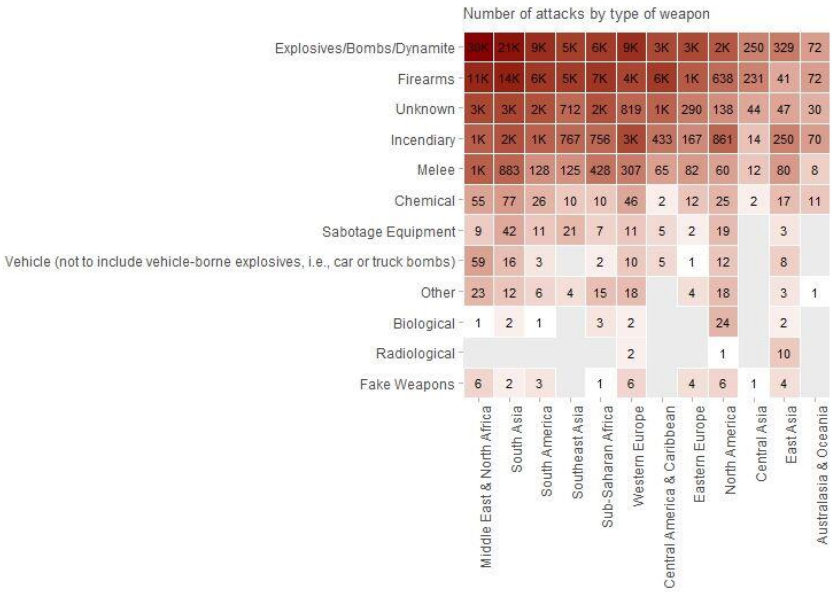
This graph shows top ten countries in Europe with highest number of attacks between 1970-2016. According to this graph, United Kingdom is coming first in Europe fort he number of attacks. This number has reached more than 5000. Spain comes in second in Europe fort he number of attacks. France, Russia Ukraine and Italy also have high number of attacks among the other european countries. Ireland is the last country in Europe fort he number of attacks. The number of attacks of Ireland is almost negligible.

Worldwide Attacks with Kills between 2000-2016



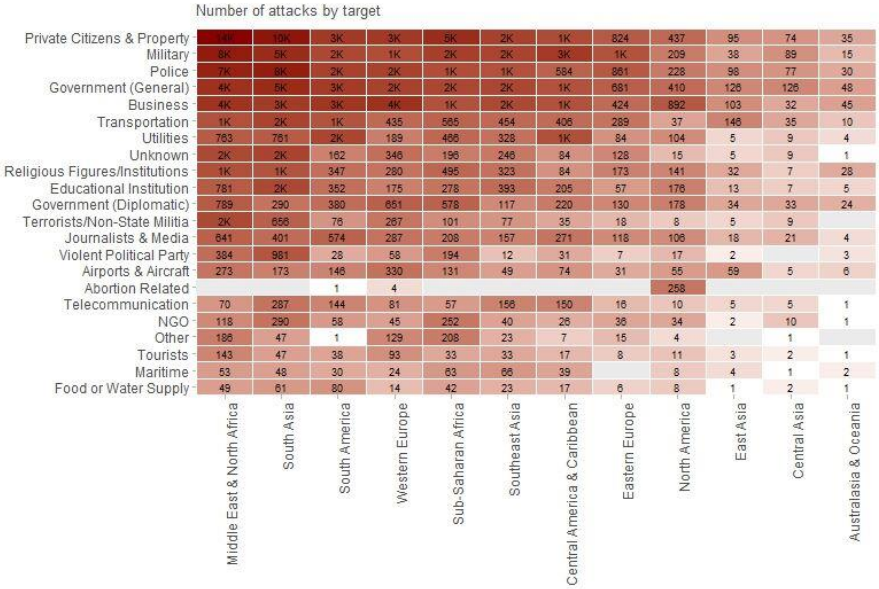
This photograph shows worldwide attacks with kills between 2000-2016. Red regions Show the highest number of attacks happened between 2000-2016. According to this analysis, Middle-East and South Asia is still coming first fort he number of attacks with kills.

Number of Attacks by Type of Weapon



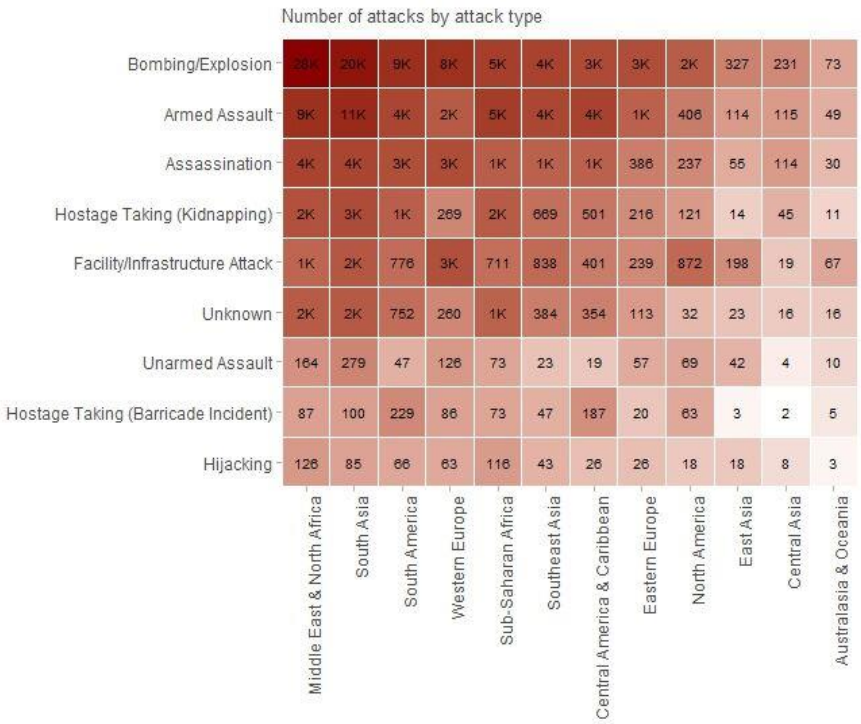
According to this tableau, expect Sab-Saharan Africa, all these regions are using explosives/ bombs/ dynamites more than the other weapons for attacks. Middle-East & North Africa is coming first for using these weapons. The number of explosives that Middle- East & North Africa used has reached 30K. Firearms come in second. Sab- Saharan Africa Region is using Firearms more than Explosives.

Number of Attacks by Target



This tableau shows number of attacks by target. According to analyses, private citizens & property is coming first for all these regions except Central America & Caribbean. For Central America & Caribbean region,number of of attacks that target military is more than the others. Middle East & North Africa region has the highest number of attacks and this number has reached 14K fort the target of pricate citizens & property.

Number of Attacks by Attack Type



This tableau shows number of attacks by attack types. According to analyses, bombing/ explosion is coming first for all the regions except Central America & Caribbeans. For Central America & Caribbeans, armed assault is in the first place. Again, Middle East & North Africa is coming first and the number of attacks that has been made by bombing / explosion has reached 23K. Armed assault is following bombing/ explosin with the number of 9K.Even the sum of all number of attacks for Central America & Caribbeans, Eastern Europe, North America, East Asia, Central Asia, Australia & Oceania is smaller than number of attacks that has been made against Middle East & North Africa by using bomb/ explosion.

3. Data Preparation

The objective of my model is to predict information for Europe.

 # Getting attacks from Europe with kills

```
training_data_europe <- dataset %>%

filter(nkill > 0) %>%

filter(region_txt == "Western Europe" | region_txt == "Eastern Europe") %>%

select(imonth, country_txt, attacktype1_txt)
```

 # Creating the models

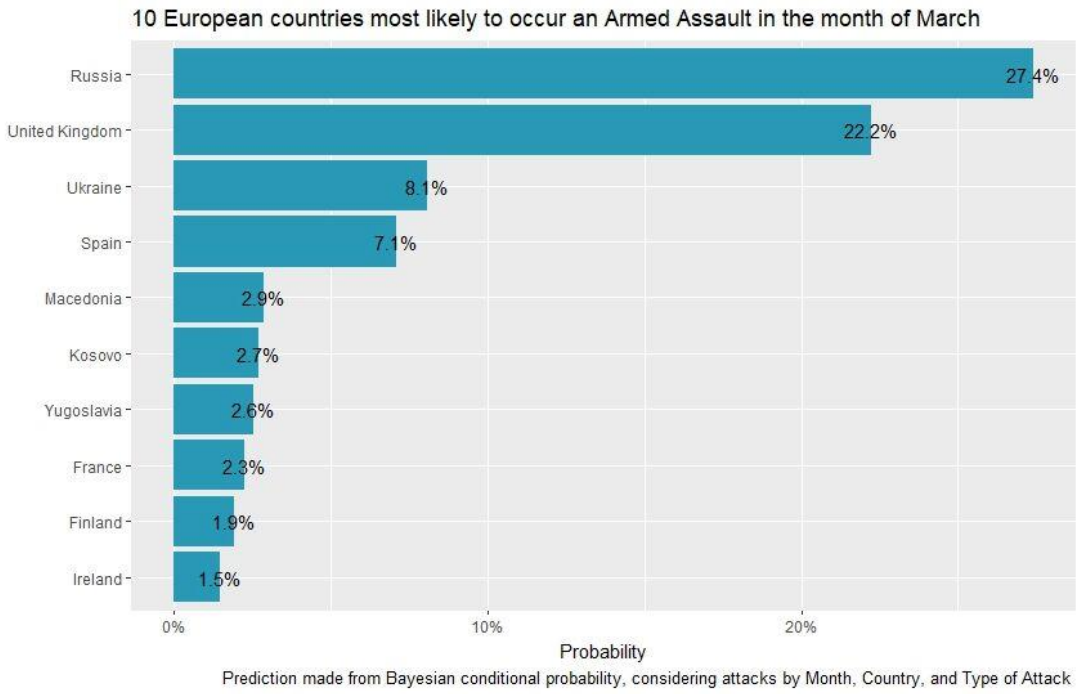
```
model_country_eur <- naiveBayes(country_txt ~., data = training_data_europe, laplace = T)     #to predic the country

model_attacktype_eur <- naiveBayes(attacktype1_txt ~., data = training_data_europe, laplace = T)   # to predic the type of attack

model_month_eur <- naiveBayes(imonth ~., data = training_data_europe, laplace = T) # to predic the month
```

That basically filter the whole dataset for just European cases and create 3 different models (one for predicting the country, another for predicting the type of attack and the last one for predicting the month).

4. Modeling

Naive Bayes Prediction Model has been used to predict the probabilities of occurence of an attack, given some circumstances. This model uses conditional probabilities and it works to answer questions like:

Considering all the attacks that already happened in Western Europe, what are the countries most likely to have an Armed Assault in March? And the algorithm returns all the countries and their respective probabilities. Like in the image below:



5. Evaluation

In the model that predicts the country given the type of attack and month, the Naive Bayes calculates the 'a-priori' probabilites, which means the probability of hapenning an attack in a country without knowing the type of the attack or the month (see image "1. a-priori.jpg")

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
          Albania           Austria           Belarus           Belgium Bosnia-Herzegovina          Bulgaria
0.0036968577 0.0024029575 0.0003696858 0.0049907579 0.0053604436 0.0027726433
          Croatia            Cyprus    Czech Republic    Czechoslovakia           Denmark East Germany (GDR)
0.0031423290 0.0036968577 0.0009242144 0.0003696858 0.0009242144 0.0003696858
          Estonia           Finland            France           Germany            Greece           Hungary
0.0005545287 0.0001848429 0.0360443623 0.0081330869 0.0123844732 0.0007393715
          Ireland             Italy            Kosovo            Latvia         Lithuania          Macedonia
0.0121996303 0.0327171904 0.0086876155 0.0003696858 0.0001848429 0.0036968577
            Malta           Moldova        Montenegro       Netherlands            Norway            Poland
0.0007393715 0.0011090573 0.0001848429 0.0033271719 0.0007393715 0.0011090573
         Portugal           Romania            Russia            Serbia  Serbia-Montenegro    Slovak Republic
0.0038817006 0.0005545287 0.1805914972 0.0003696858 0.0005545287 0.0011090573
         Slovenia      Soviet Union             Spain            Sweden       Switzerland           Ukraine
0.0001848429 0.0046210721 0.1356746765 0.0022181146 0.0020332717 0.0709796673
    United Kingdom West Germany (FRG)        Yugoslavia
0.4255083179 0.0085027726 0.0110905730
```

The Naive Bayes also calculates the conditional probabilites for the other variables. To the month, for example, the model calculates the probability of having an attack in a specific month, given the country (see image "2. conditional probabilities.jpg"; example: the probability of having an attack in January given the country is Albania, is 0.1 (10%)).

Conditional probabilities:

Y	imonth 1	10	11	12	2	3	4	5	6
Albania	0.10000000	0.05000000	0.10000000	0.10000000	0.10000000	0.15000000	0.05000000	0.05000000	0.45000000
Austria	0.23076923	0.15384615	0.15384615	0.23076923	0.15384615	0.07692308	0.15384615	0.23076923	0.15384615
Belarus	0.50000000	1.00000000	0.50000000	0.50000000	0.50000000	0.50000000	1.00000000	0.50000000	0.50000000
Belgium	0.03703704	0.14814815	0.07407407	0.11111111	0.07407407	0.37037037	0.03703704	0.11111111	0.11111111
Bosnia-Herzegovina	0.03448276	0.17241379	0.13793103	0.03448276	0.06896552	0.10344828	0.13793103	0.10344828	0.10344828
Bulgaria	0.13333333	0.40000000	0.20000000	0.06666667	0.06666667	0.06666667	0.06666667	0.20000000	0.13333333
Croatia	0.11764706	0.11764706	0.11764706	0.11764706	0.05882353	0.11764706	0.11764706	0.11764706	0.05882353
Cyprus	0.05000000	0.05000000	0.15000000	0.10000000	0.15000000	0.05000000	0.20000000	0.20000000	0.15000000
Czech Republic	0.60000000	0.20000000	0.60000000	0.20000000	0.20000000	0.20000000	0.20000000	0.40000000	0.20000000
Czechoslovakia	1.00000000	0.50000000	0.50000000	0.50000000	0.50000000	0.50000000	0.50000000	0.50000000	0.50000000
Denmark	0.20000000	0.20000000	0.20000000	0.20000000	0.60000000	0.40000000	0.20000000	0.20000000	0.20000000
East Germany (GDR)	0.50000000	0.50000000	1.00000000	0.50000000	0.50000000	0.50000000	0.50000000	1.00000000	0.50000000
Estonia	0.33333333	0.33333333	0.33333333	0.33333333	0.66666667	0.66666667	0.33333333	0.33333333	0.33333333
Finland	1.00000000	1.00000000	2.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
France	0.10256410	0.06666667	0.07179487	0.13846154	0.09743590	0.10256410	0.06153846	0.06153846	0.08205128
Germany	0.04545455	0.09090909	0.13636364	0.09090909	0.09090909	0.06818182	0.13636364	0.09090909	0.13636364
Greece	0.10447761	0.07462687	0.14925373	0.07462687	0.05970149	0.13432836	0.10447761	0.05970149	0.11940299
Hungary	0.50000000	0.25000000	0.25000000	0.25000000	0.50000000	0.25000000	0.25000000	0.25000000	0.50000000
Ireland	0.15151515	0.12121212	0.04545455	0.12121212	0.10606061	0.09090909	0.04545455	0.15151515	0.12121212
Italy	0.10169492	0.07344633	0.06779661	0.13559322	0.06779661	0.14124294	0.10734463	0.08474576	0.11864407
Kosovo	0.12765957	0.04255319	0.06382979	0.06382979	0.08510638	0.02127660	0.10638298	0.04255319	0.23404255
Latvia	0.50000000	0.50000000	0.50000000	0.50000000	0.50000000	1.00000000	0.50000000	0.50000000	0.50000000
Lithuania	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	2.00000000	1.00000000
Macedonia	0.20000000	0.10000000	0.10000000	0.05000000	0.10000000	0.10000000	0.15000000	0.20000000	0.10000000
Malta	0.25000000	0.50000000	0.25000000	0.75000000	0.25000000	0.50000000	0.25000000	0.25000000	0.25000000
Moldova	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667	0.50000000	0.33333333	0.16666667
Montenegro	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	2.00000000	1.00000000	1.00000000	1.00000000
Netherlands	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	0.16666667	0.05555556	0.38888889	0.16666667
Norway	0.50000000	0.25000000	0.50000000	0.25000000	0.25000000	0.25000000	0.25000000	0.25000000	0.25000000
Poland	0.16666667	0.33333333	0.16666667	0.16666667	0.16666667	0.16666667	0.33333333	0.50000000	0.16666667
Portugal	0.04761905	0.23809524	0.14285714	0.09523810	0.09523810	0.04761905	0.23809524	0.19047619	0.14285714
Romania	0.33333333	0.33333333	0.33333333	0.66666667	0.33333333	0.33333333	0.33333333	0.66666667	0.33333333
Russia	0.07471853	0.08290686	0.09211873	0.07471853	0.05936540	0.05117707	0.06960082	0.08290686	0.08188332
Serbia	0.50000000	0.50000000	0.50000000	0.50000000	1.00000000	0.50000000	0.50000000	0.50000000	0.50000000
Serbia-Montenegro	0.33333333	0.33333333	0.33333333	0.33333333	0.66666667	1.00000000	0.33333333	0.33333333	0.33333333
Slovak Republic	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667	0.50000000	0.33333333	0.16666667
Slovenia	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
Soviet Union	0.08000000	0.12000000	0.04000000	0.04000000	0.04000000	0.24000000	0.08000000	0.12000000	0.12000000
Spain	0.08583106	0.11035422	0.08855586	0.07629428	0.06539510	0.09264305	0.07629428	0.09264305	0.09945504
Sweden	0.16666667	0.16666667	0.16666667	0.25000000	0.33333333	0.08333333	0.25000000	0.08333333	0.08333333
Switzerland	0.09090909	0.09090909	0.36363636	0.09090909	0.18181818	0.27272727	0.18181818	0.09090909	0.18181818
Ukraine	0.09375000	0.11718750	0.10156250	0.03385417	0.09895833	0.03906250	0.04687500	0.12239583	0.07031250
United Kingdom	0.07428323	0.09513467	0.09774109	0.06342311	0.07775847	0.08861859	0.08948740	0.08253692	0.08340573
West Germany (FRG)	0.10869565	0.08695652	0.08695652	0.06521739	0.15217391	0.04347826	0.08695652	0.13043478	0.15217391
Yugoslavia	0.08333333	0.10000000	0.10000000	0.03333333	0.08333333	0.20000000	0.11666667	0.10000000	0.15000000

The conditional probability works in the same way for the attack type given the month (image "3. conditional probability").

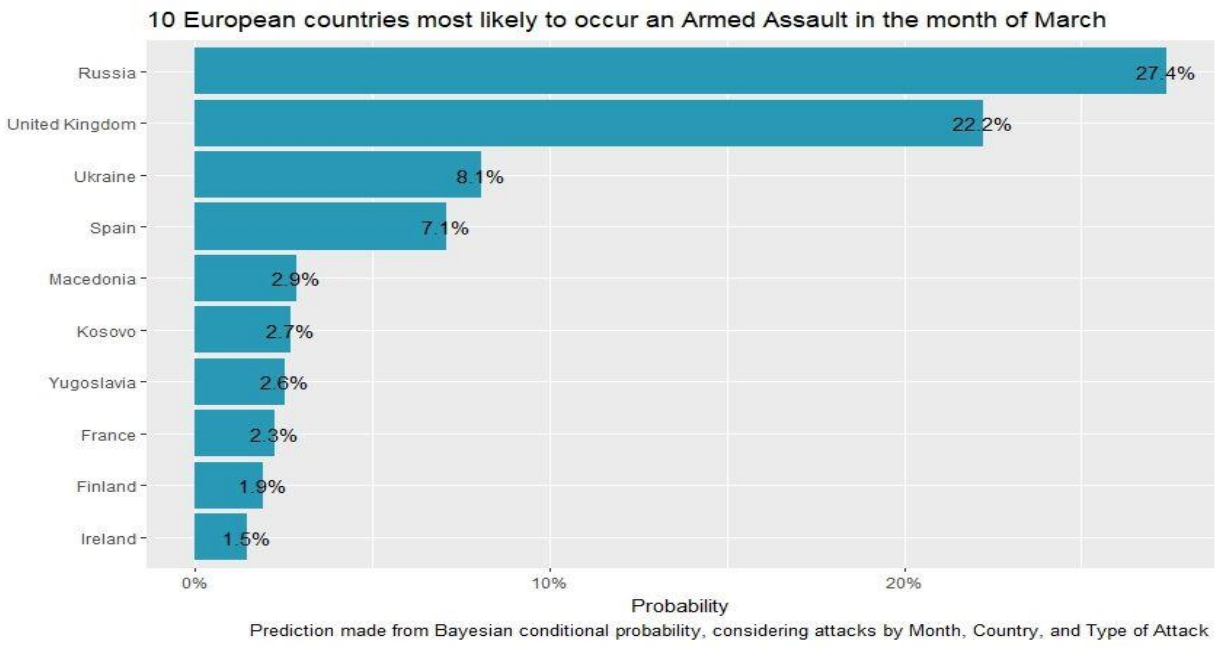
Y	attacktype1_txt				
	Armed Assault	Assassination	Bombing/Explosion	Facility/Infrastructure Attack	Hijacking
Albania	0.350000000	0.400000000	0.300000000	0.050000000	0.050000000
Austria	0.153846154	0.461538462	0.461538462	0.076923077	0.153846154
Belarus	0.500000000	0.500000000	1.500000000	0.500000000	0.500000000
Belgium	0.222222222	0.555555556	0.259259259	0.074074074	0.037037037
Bosnia-Herzegovina	0.379310345	0.241379310	0.482758621	0.034482759	0.034482759
Bulgaria	0.066666667	0.466666667	0.666666667	0.066666667	0.066666667
Croatia	0.529411765	0.117647059	0.529411765	0.058823529	0.058823529
Cyprus	0.150000000	0.650000000	0.250000000	0.050000000	0.150000000
Czech Republic	0.200000000	0.400000000	0.800000000	0.200000000	0.200000000
Czechoslovakia	0.500000000	0.500000000	1.000000000	0.500000000	0.500000000
Denmark	0.800000000	0.200000000	0.600000000	0.200000000	0.200000000
East Germany (GDR)	0.500000000	1.000000000	0.500000000	0.500000000	0.500000000
Estonia	0.333333333	0.666666667	1.000000000	0.333333333	0.333333333
Finland	2.000000000	1.000000000	1.000000000	1.000000000	1.000000000
France	0.117948718	0.543589744	0.271794872	0.005128205	0.005128205
Germany	0.522727273	0.272727273	0.045454545	0.113636364	0.068181818
Greece	0.119402985	0.522388060	0.313432836	0.044776119	0.059701493
Hungary	0.250000000	0.250000000	0.750000000	0.500000000	0.250000000
Ireland	0.151515152	0.575757576	0.181818182	0.030303030	0.015151515
Italy	0.079096045	0.745762712	0.118644068	0.033898305	0.011299435
Kosovo	0.468085106	0.127659574	0.425531915	0.021276596	0.021276596
Latvia	0.500000000	1.000000000	1.000000000	0.500000000	0.500000000
Lithuania	1.000000000	1.000000000	2.000000000	1.000000000	1.000000000
Macedonia	0.750000000	0.100000000	0.200000000	0.050000000	0.050000000
Malta	0.250000000	1.000000000	0.500000000	0.250000000	0.250000000
Moldova	0.500000000	0.166666667	0.833333333	0.166666667	0.166666667
Montenegro	1.000000000	1.000000000	2.000000000	1.000000000	1.000000000
Netherlands	0.388888889	0.555555556	0.111111111	0.055555556	0.111111111
Norway	0.750000000	0.250000000	0.750000000	0.250000000	0.250000000
Poland	0.166666667	0.833333333	0.500000000	0.166666667	0.166666667
Portugal	0.095238095	0.571428571	0.428571429	0.047619048	0.047619048
Romania	0.333333333	1.000000000	0.666666667	0.333333333	0.333333333
Russia	0.390992835	0.184237462	0.371545548	0.009211873	0.008188332
Serbia	1.000000000	0.500000000	1.000000000	0.500000000	0.500000000
Serbia-Montenegro	0.666666667	1.000000000	0.333333333	0.333333333	0.333333333
Slovak Republic	0.333333333	0.166666667	1.000000000	0.166666667	0.166666667
Slovenia	1.000000000	2.000000000	1.000000000	1.000000000	1.000000000
Soviet Union	0.520000000	0.240000000	0.360000000	0.040000000	0.040000000
Spain	0.117166213	0.640326975	0.217983651	0.008174387	0.002724796
Sweden	0.333333333	0.500000000	0.250000000	0.083333333	0.083333333
Switzerland	0.272727273	0.545454545	0.363636364	0.090909091	0.181818182
Ukraine	0.234375000	0.049479167	0.609375000	0.007812500	0.002604167
United Kingdom	0.135534318	0.622936577	0.210686360	0.003040834	0.001303215
West Germany (FRG)	0.108695652	0.521739130	0.347826087	0.043478261	0.043478261
Yugoslavia	0.533333333	0.250000000	0.216666667	0.016666667	0.016666667

Then, when we want to predict something, the algorithm multiplicates all these probabilites for classificating the most likely results.

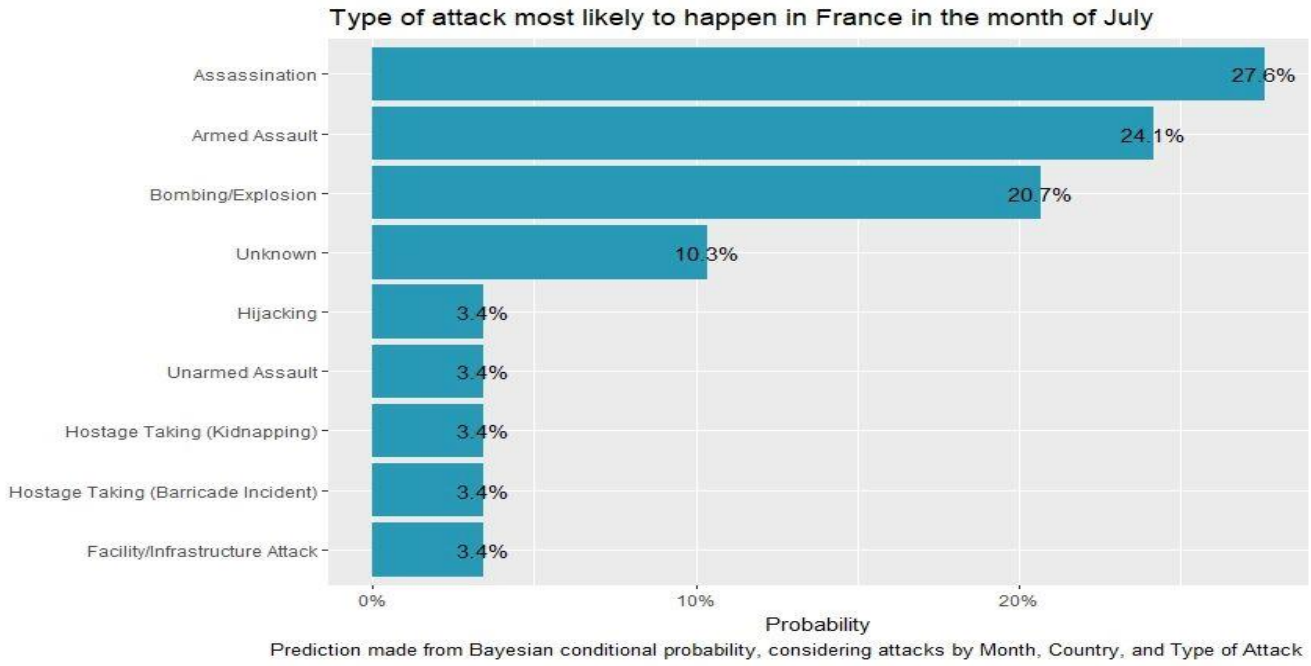
For example, when we want to predict the country, we give as parameters the type of attack and month. Then the algorithm multiplicates all the conditional probabilities and the 'a-priori' probabilities to result the most likely countries to suffer that type of attack in that month we specified (See image "4. Result of the multiplication of the probabilities.jpg")

```
        Albania     Austria     Belarus     Belgium Bosnia-Herzegovina    Bulgaria     Croatia      Cyprus Czech Republic
[1,] 0.006719361 0.004430348 0.004799544 0.00213313        0.003641033 0.001279878 0.01016374 0.001439863    0.005759452
     Czechoslovakia     Denmark East Germany (GDR)      Estonia     Finland     France     Germany      Greece     Hungary
[1,]    0.009599087 0.00767927        0.004799544 0.003199696 0.01919817 0.022644 0.01003541 0.008023118 0.004799544
        Ireland      Italy      Kosovo      Latvia   Lithuania  Macedonia       Malta     Moldova  Montenegro Netherlands
[1,] 0.01454407 0.0136665 0.02695914 0.004799544 0.009599087 0.02879726 0.002399772 0.004799544 0.009599087 0.007465957
        Norway      Poland     Portugal     Romania      Russia      Serbia Serbia-Montenegro Slovak Republic    Slovenia
[1,] 0.01439863 0.001599848 0.0009141988 0.003199696 0.2739817 0.009599087       0.006399391     0.003199696 0.009599087
     Soviet Union      Spain      Sweden Switzerland     Ukraine United Kingdom West Germany (FRG) Yugoslavia
[1,]  0.009983051 0.07085539 0.006399391 0.002617933 0.0809923       0.222472        0.005216895 0.02559757
```

6. Deployment



10 European countries most likely to occur an Armed Assault in March the results means that considering all the past armed assaults that happened in Europe in March, those are the countries most likely to suffer.



This result shows the type of attack most likely to happen in France in the month of July. According to the result, assassination is coming first with 27.6% percentage.